19 RÉPUBLIQUE FRANÇAISE

INSTITUT NATIONAL DE LA PROPRIÉTÉ INDUSTRIELLE

PARIS

11) No de publication :

2 822 999

(à n'utiliser que pour les commandes de reproduction)

②1) No d'enregistrement national :

01 04194



(51) Int Cl⁷: **G 10 L 21/02**, H 04 M 3/40, 9/00, 1/253, H 04 L 27/01, H 03 G 3/20, 5/22

DEMANDE DE BREVET D'INVENTION

Α1

22) Date de dépôt : 28.03.01.

(30) Priorité :

71) Demandeur(s): FRANCE TELECOM Société anonyme — FR.

Date de mise à la disposition du public de la demande : 04.10.02 Bulletin 02/40.

(56) Liste des documents cités dans le rapport de recherche préliminaire : Se reporter à la fin du présent fascicule

Références à d'autres documents nationaux apparentés :

(72) Inventeur(s): MAHE GAEL et GILLOIRE ANDRE.

73 Titulaire(s) :

(74) Mandataire(s): CABINET BALLOT.

PROCEDE ET DISPOSITIF DE CORRECTION CENTRALISEE DU TIMBRE DE LA PAROLE SUR UN RESEAU DE COMMUNICATIONS TELEPHONIQUES.

(57) L'invention concerne un procédé et dispositif de correction du timbre de la parole transmise sous la forme d'un signal au moyen d'une liaison de transmission d'un réseau téléphonique.

La correction ayant lieu après la conversion analogiquenumérique du signal dans le réseau, et comprend une préégalisation (23) du signal numérique par un filtre fixe ayant une réponse fréquentielle dans une bande de fréquences l'c-Fh correspondant à l'inverse d'une déformation spectrale de référence introduite par la liaison téléphonique, avec Fc < 300Hz et Fh ≥ 3150Hz et une égalisation adaptée (24) dans laquelle on utilise un filtre ayant une réponse fréquentielle adaptée automatiquement à la distorsion réelle intro-

tielle adaptée automatiquement à la distorsion réelle introduite par la liaison téléphonique en fonction du rapport entre un spectre de référence et un spectre correspondant au spectre à long terme du signal. Pré-égalisateur Egalisateur adapté CAG CAG

FN 2 822 999 - A1



PROCEDE ET DISPOSITIF DE CORRECTION CENTRALISEE DU TIMBRE DE LA PAROLE SUR UN RESEAU DE COMMUNICATIONS TELEPHONIQUES

L'invention a pour objet un procédé et un dispositif de correction centralisée du timbre de la parole sur un réseau de communications téléphoniques.

L'invention s'applique à tout type de réseau de communication (fixe, mobile ou autre) introduisant des déformations spectrales et des modifications du niveau du signal, moyennant le choix approprié de certains paramètres.

5

10

15

20

Dans le cas d'un réseau téléphonique filaire, la parole subit deux distorsions spectrales.

La première distorsion est le filtrage passe-bande (300-3400Hz) aux extrémités de chaque ligne analogique d'abonné (terminal téléphonique émetteur - respectivement récepteur - et point de conversion analogique-numérique - respectivement numérique-analogique), normalisé sous le nom de « Système de Référence Intermédiaire » (SRI) que l'on trouve défini par la recommandation p48 de l'UIT.

On pourra se reporter aux figures la et 1b représentant les réponses fréquentielles des systèmes d'émission et de réception définis par l'UIT.

Ce filtrage passe-bande dégrade le timbre en atténuant fortement les composantes basse-fréquence de la parole.

La deuxième distorsion est celle introduite par les lignes analogiques elles-mêmes, qui constituent des filtres passe-bas dont la pente est d'autant plus raide que la ligne est longue. Dans un modèle simple de ligne

analogique, l'atténuation en dB est proportionnelle à la racine carrée de la fréquence :

$$H_{dB}(f) = H_{dB}(800 \text{ Hz}) \sqrt{\frac{f}{800}}$$
 (1)

5

10

15

20

avec $H_{dB}(800 Hz)$, valant 3 dB pour les lignes moyennes et 9,5 dB pour les lignes les plus longues.

Les réponses fréquentielles de différentes lignes (courte, moyenne et longue) sont représentées sur la figure 2.

Ce filtrage passe-bas a pour effet d'assourdir la voix du locuteur.

Dans le cas d'un réseau mobile, le signal subit uniquement un filtrage passe-bande 300-3400Hz au niveau des terminaux émetteur et récepteur. Ce filtrage passe bande doit respecter un gabarit défini par la recommandation P.313 de l'UIT.

Jusqu'à présent la compensation des distorsions spectrales introduites dans le signal de parole par les divers éléments de la liaison téléphonique est réalisée par des dispositifs à base d'égalisation. Celle-ci peut être fixe ou s'adapter en fonction des conditions de transmission.

Un premier état de la technique concerne les dispositifs d'égalisation fixe centralisée.

En effet, des dispositifs d'égalisation centralisée ont été proposés dans les brevets US 5333195 et US 5471527.

Ces égaliseurs sont des filtres fixes qui restaurent le niveau des basses fréquences atténuées par l'émetteur. Le brevet US 5333195 propose par exemple un gain de 10 à 15 dB sur la bande 100-300Hz. Ces méthodes présentent deux inconvénients :

- l'égaliseur ne compense que le filtrage de l'émetteur, de sorte qu'à la réception les composantes basse-fréquence restent fortement affaiblies par le filtrage SRI de réception.
- Cette égalisation fixe compense des conditions de transmission (ligne et système d'émission) moyennes. Si les conditions réelles sont trop différentes (par exemple si les lignes analogiques sont lonques) le dispositif corrige pas suffisamment le timbre, l'altère plus que la liaison sans égalisation.

5

10

25

Un deuxième état de la technique repose sur des dispositifs à égalisation adaptative.

Le dispositif décrit dans le brevet US 5915235 vise 15 à corriger la réponse fréquentielle non idéale d'un transducteur de téléphone mobile. L'éqaliseur décrit comme étant placé entre convertisseur un analogique-numérique et un codeur CELP (Code Exited Linear Predictive Coding), mais peut être aussi bien 20 dans le terminal téléphonique que dans le réseau.

Deux méthodes sont alors proposées dans ce document :

 La première consiste à calculer les coefficients d'auto-corrélation à long terme R_{LT}:

$$R_{LT}(n,i) = \alpha R_{LT}(n-1,i) + (1-\alpha)R(n,i),$$
 (2)

Avec : $R_{LT}(n,i)$ i^{ième} coefficient d'auto-corrélation à long terme à la n^{ième} trame, R(n,i) i^{ième} coefficient d'auto-corrélation spécifique à la n^{ième} trame, et α constante de lissage fixée par exemple à 0,995. De ces coefficients sont dérivés les coefficients LPC (Linear Predictive Coding) à long terme, qui sont les

coefficients d'un filtre blanchisseur. A la sortie de ce filtre, le signal est filtré par un filtre fixe qui lui imprime les caractéristiques spectrales à long terme idéales, c'est-à-dire celles qu'il aurait à la sortie d'un transducteur ayant la réponse fréquentielle idéale. Ces deux filtres sont complétés par un gain multiplicatif égal au rapport entre les énergies à long terme de l'entrée du filtre blanchisseur et la sortie du deuxième filtre.

- La deuxième méthode consiste à diviser le signal en sous-bandes et, pour chaque sous-bande, appliquer un gain multiplicatif de manière à atteindre une énergie cible, ce gain étant défini comme le rapport entre l'énergie cible de la sous-bande et l'énergie à long terme (obtenue par un lissage de l'énergie instantanée) du signal dans cette sous-bande.

Ces deux méthodes présentent l'inconvénient de ne corriger que la réponse non idéale du système d'émission et non celle du système de réception.

20

25

- 30

5

Le dispositif décrit dans le brevet FR 9408741 (US 5905969) a pour objet de compenser le filtrage du système d'émission et de la ligne d'abonné d'émission pour améliorer la reconnaissance centralisée de parole et/ou la qualité de la parole transmise. Comme le représente la figure 3a de ce brevet, le spectre du signal est divisé en 24 sous-bandes et chaque énergie de sous-bande est multipliée par un gain adaptatif. Une adaptation du gain est réalisée selon l'algorithme du gradient stochastique, par minimisation de quadratique, l'erreur étant définie comme la différence entre l'énergie de sous-bande et une énergie référence définie pour chaque sous-bande. L'énergie de référence est modulée à chaque trame par l'énergie

globale de la trame courante, de manière à respecter les variations naturelles de niveau à court terme du signal de parole. La convergence de l'algorithme permet d'obtenir en sortie les 24 signaux de sous-bande égalisés.

Le dispositif ne corrige pas le filtrage du système de réception et de la ligne analogique de réception.

5

10

15

20

25

30

Ce brevet ne mentionne pas de résultats en termes d'amélioration de la qualité vocale et reconnaît que la méthode est sous-optimale, car on opère une transformation dans le domaine fréquentiel avec une erreur due à l'opération de convolution circulaire implicite liée à cette transformation.

En outre, il y a une oscillation du système autour de la solution optimale liée au fait que la correction est réalisée par un algorithme adaptatif en boucle fermée (boucle de rétroaction).

Une compensation de l'effet de ligne est décrite dans l'article « On line adaptation of a speech recognizer to variation in telephone lines conditions », Eurospeech, pp 1247-1250, sept.1993 de C.Mokbel, J.Monné and D. Jouvet par la méthode de la soustraction cepstrale, dans le but d'améliorer la robustesse de la reconnaissance de la parole.

On montre dans ce document que le cepstre du canal de transmission peut être estimé par le cepstre moyen du signal reçu, celui-ci étant préalablement blanchi par un filtre de pré-accentuation.

Cette méthode permet une nette amélioration des performances des systèmes de reconnaissance de la parole, mais est considérée comme une méthode « offline », 2 à 4 secondes étant nécessaires pour estimer le cepstre moyen. Elle ne peut pas par conséquent

s'appliquer à la correction des distorsions sur la parole introduites par le canal de transmission d'un réseau téléphonique.

La présente invention a pour but l'amélioration de la qualité de la parole transmise sur les réseaux de communication, en offrant des moyens pour corriger les déformations spectrales du signal de parole et écarts de niveau rapport au par niveau nominal souhaitable pour la perception de la parole, déformations et écarts provoqués par différents maillons de la chaîne de transmission.

5

10

15

20

25

30

L'invention a plus particulièrement pour objet un procédé de correction du timbre de la parole transmise sous la forme d'un signal au moyen d'une liaison de transmission d'un réseau téléphonique, la correction ayant lieu après la conversion analogique-numérique du signal dans le réseau, principalement caractérisé en ce qu'il comprend au moins une étape de pré-égalisation du signal numérique par un filtre fixe ayant une réponse fréquentielle dans une bande de fréquences correspondant à l'inverse d'une déformation spectrale de référence introduite par la liaison de transmission, avec Fc < 300Hz et Fh ≥ 3150 Hz.

Selon une autre caractéristique, la déformation spectrale prise comme référence est caractérisée, dans le cas d'une liaison d'un réseau de téléphonie commutée (RTC), à partir de la mise en cascade d'un système de référence intermédiaire (SRI) du type défini par la recommandation UIT-T P.48, et de deux lignes analogiques moyennes (émission et réception).

Dans le cas où le terminal émetteur est un téléphone mobile et le terminal récepteur est un poste

fixe, la déformation spectrale prise comme référence est celle résultant de la mise en cascade d'un filtre respectant le gabarit d'efficacité des mobiles en émission défini par la recommandation P.313 de l'UIT, d'une ligne analogique moyenne et d'un système récepteur du SRI tel que défini par la recommandation UIT-TP.48.

5

10

15

20

25

30

Dans le cas où le terminal émetteur est un poste fixe et le terminal récepteur est un téléphone mobile, la réponse spectrale prise comme référence est celle résultant de la mise en cascade d'un système émetteur du SRI tel que défini par la recommandation UIT-TP.48, d'une ligne analogique moyenne et d'un filtre respectant le gabarit d'efficacité des mobiles en réception défini par la recommandation P.313 de l'UIT.

Dans le cas d'une liaison entre terminaux mobiles, la déformation spectrale prise comme référence est, celle résultant de la mise en cascade d'un filtre respectant le gabarit d'efficacité des mobiles en émission défini par la recommandation P.313 de l'UIT et d'un filtre respectant le gabarit d'efficacité des mobiles en réception défini par la recommandation P.313 de l'UIT.

Selon une autre caractéristique, le procédé de correction du timbre de la parole comporte en outre une étape d'égalisation adaptée dans laquelle on utilise un filtre ayant une réponse fréquentielle adaptée automatiquement à la distorsion réelle introduite par la liaison téléphonique en fonction du rapport entre un spectre de référence et un spectre correspondant au spectre à long terme du signal.

Selon une autre caractéristique, l'étape d'égalisation adaptée comprend :

- la détection d'une activité vocale sur la ligne pour déclencher un enchaînement de traitements pour le calcul des coefficients du filtre numérique en fonction du rapport entre le spectre de référence et le spectre correspondant au spectre à long terme du signal,
- la commande du filtre avec les coefficients obtenus et l'actualisation desdits coefficients,
- le filtrage du signal sortant du pré-égaliseur par ledit filtre.
- Selon une autre caractéristique, l'enchaînement de traitements comprend :

5

15

20

25

- 30

- le calcul du spectre à long terme du signal dans des fenêtres temporelles successives se recouvrant partiellement, ce calcul étant réalisé dans la bande de fréquences Fc-Fh,
- le calcul du module de la réponse fréquentielle de l'égaliseur adapté en effectuant le rapport de la racine carrée du spectre à long terme obtenu dans une fenêtre temporelle, à la racine carrée du spectre de référence, la racine carrée du spectre de référence étant compensé à chaque fréquence par un facteur prédéterminé A(f) fonction de la fréquence.

Selon une autre caractéristique, l'enchaînement de traitements comprend une extrapolation du module de la réponse fréquentielle de l'égaliseur adapté, pour les fréquences en dehors de la bande Fc-Fh, la réponse fréquentielle étant définie pour toutes les fréquences comprises entre 0-4000Hz.

L'enchaînement de traitements comprend en outre le calcul de la réponse impulsionnelle du filtre numérique à partir du module de la réponse fréquentielle de l'égaliseur adapté extrapolé pour les fréquences en dehors de la bande Fc-Fh.

A Comment

Le calcul du spectre du signal à long terme comprend une opération de transformée de Fourier rapide.

Le calcul de la réponse impulsionnelle du filtre consiste à calculer les coefficients du filtre en opérant une transformée de Fourier inverse sur le module de la réponse fréquentielle, suivie d'une symétrisation, d'un fenêtrage et d'un décalage.

L'application d'une fenêtre temporelle correspond à un lissage de la réponse fréquentielle initiale calculée.

Le calcul du spectre à long terme du signal dans des fenêtres temporelles successives se recouvrant partiellement comprend :

- l'échantillonnage du signal dans une fenêtre temporelle,
- l'opération de transformée de Fourier rapide (FFT) du signal échantillonné,
- le calcul de la densité spectrale de puissance,
- le calcul de la moyenne de la densité spectrale de puissance, sur une durée prédéterminée.

Avantageusement, le calcul de la moyenne de la densité spectrale de puissance consiste :

- pour les N premières fenêtres temporelles à partir de la détection de présence de parole calculer la. moyenne dans le signal, à spectrales des densités arithmétique puissance de toutes les fenêtres temporelles écoulées depuis ladite détection, N étant un temporelles prédéterminé, fenêtre nombre de typiquement mais non exclusivement le nombre de fenêtres temporelles dans 4 secondes de parole;
- pour les fenêtres temporelles suivantes, à ajuster la moyenne de la densité spectrale de

15

5

10

20

25

puissance calculée à la fenêtre temporelle précédente par un lissage récursif du premier ordre tenant compte de la densité spectrale de puissance de la fenêtre temporelle courante.

Cela se traduït par la formule générique suivante :

5

20

25

-30

$$E[\gamma_x(f)]_n = \alpha(n)\gamma_x(f,n) + (1-\alpha(n))E[\gamma_x(f)]_{n-1}, \qquad (9)$$

où $E[\gamma_x(f)]_n$ est le spectre à long terme de x à la $n^{\text{ième}}$ 10 trame $\gamma_x(f,n)$ la densité spectrale de puissance de la nième trame, et

$$\alpha(n) = \frac{1}{\min(n, N)}$$

Le procédé comprend en outre une étape de contrôle automatique du gain.

Selon un mode de réalisation, le contrôle automatique du gain est réalisé durant l'enchaînement des traitements de l'étape d'égalisation en choisissant une densité spectrale de référence γ_{ref} correspondant au niveau souhaité en réception.

Selon un autre mode de réalisation, le contrôle automatique du gain est réalisé par amplification du signal obtenu après égalisation avec un gain α fonction du rapport entre la densité spectrale de puissance du signal de sortie du terminal de réception lorsque l'on a effectué une pré-égalisation et une égalisation adaptée du signal et, la densité spectrale de puissance du signal de sortie du terminal de réception en l'absence de pré-égalisation et d'égalisation adaptée du signal.

Un autre objet de l'invention est un filtre numérique fixe destiné à la correction du timbre de la parole dans un réseau de transmission téléphonique,

principalement caractérisé en ce que ledit filtre a une réponse fréquentielle dans une bande de fréquences Fc-Fh, correspondant à l'inverse d'une déformation spectrale de référence introduite par la liaison, avec Fc < 300Hz et Fh ≥ 3150Hz.

5

10

15

20

25

30

Un autre objet de l'invention est numérique adapté destiné à la correction du timbre de la parole dans un réseau de transmission téléphonique, principalement caractérisé en ce qu'il comprend des moyens de traitement du signal de parole ayant une réponse fréquentielle adaptée automatiquement distorsion réelle introduite par la liaison téléphonique en fonction du rapport entre un spectre de référence et un spectre correspondant au spectre à long terme du signal.

Un autre objet de l'invention est un dispositif de correction du timbre de la parole dans un réseau de transmission téléphonique, principalement caractérisé en ce qu'il comprend un filtre fixe suivi d'un filtre adapté et des moyens de contrôle automatique du gain tels que décrits précédemment.

D'autres particularités et avantages de l'invention apparaîtront clairement à la lecture de la description qui est faite ci-après et qui est donnée à titre d'exemple non limitatif et en regard des dessins sur lesquels :

- la figure la représente le gabarit du système d'émission,
- la figure lb représente le gabarit du système de réception,
 - la figure 2 représente les réponses de différentes lignes d'abonnés analogiques,

- la figure 3 représente une liaison téléphonique simplifiée incluant la correction,
- la figure 4 représente le schéma fonctionnel d'un dispositif de correction,
- La figure 5 représente la réponse fréquentielle du pré-égaliseur pour Fc=250Hz,

5

10

25

30

- La figure 6 représente une fenêtre triangulaire appliquée à la réponse impulsionnelle du filtre,
- La figure 7 représente le schéma fonctionnel de l'égaliseur adapté,
- La figure 8 représente la distorsion spectrale entre le signal émis et le signal reçu pour un locuteur 1 et pour un locuteur 2.

La description qui en est donnée dans la suite fait explicitement référence à la transmission de la parole sur lignes téléphoniques « classiques » (c'est-à-dire filaires), mais bien entendu comme cela a été dit l'invention s'applique à tout type de réseau de communication (fixe, mobile ou autre) introduisant des déformations spectrales et des modifications du niveau du signal, moyennant le choix approprié de certains paramètres.

L'objet de l'invention est de corriger distorsions spectrales par un traitement centralisé, c'est-à-dire par un dispositif 20 installé dans partie numérique du réseau téléphonique entre convertisseur analogique-numérique (loi A dans le cas particulier du RTC européen ou loi µ aux Etats-Unis) 12 et numérique (loi A)-analogique 32, comme illustré par le schéma de la figure 3. Le dispositif de correction 20 est précédé par un module 21 de conversion du signal numérique (la loi A) en linéaire et suivi par un module de conversion du signal linéaire en loi A.

La figure 4 illustre le dispositif de correction selon l'invention.

Une correction satisfaisante des distorsions moyennes dues au système émission 10, réception 30 et aux lignes analogiques 11, 31 est obtenue par un préégaliseur 23.

5

10

15

20

25

30

Afin de tenir compte du fait que les conditions de transmission ne sont pas toujours des moyennes de transmission, (les lignes utilisées ne sont longueur moyenne et toujours de les systèmes d'émission de réception peuvent s'écarter et recommandations de l'UIT), le dispositif de correction comporte en outre un égaliseur adapté 24 et une correction automatique de gain (CAG) 25. Comme on le verra dans la suite le contrôle automatique du gain peut être soit intégré à l'égaliseur adapté, soit faire l'objet d'un module séparé.

Le pré-égaliseur 23 est un filtre fixe dont la réponse fréquentielle sur une bande Fc-Fh, telle que Fc < 300Hz et Fh ≥ 3150Hz, est l'inverse de la réponse globale du canal analogique moyen d'une liaison téléphonique. Ce canal moyen est défini comme étant constitué de deux lignes d'abonné moyennes et d'un système d'émission et de réception respectant les réponses fréquentielles nominales définies dans les recommandations de l'UIT.

Fc est la fréquence basse limite d'égalisation. Elle doit être inférieure à 300 Hz de manière à restaurer les composantes basse-fréquence (BF) de la voix.

Fh est par exemple à 3150Hz.

La figure 5 représente la réponse fréquentielle typique du pré-égaliseur pour Fc=250 Hz. Cette réponse

est calculée à partir des modèles du SRI et de la « ligne moyenne ».

Le pré-égaliseur 23 ayant la réponse fréquentielle représentée sur la figure 5 est réalisé par exemple par un filtre à réponse impulsionnelle infinie IIR, dont les coefficients de la fonction de transfert en z sont :

1	0
-	~

5

20

25

30

Comme on vient de le voir, le pré-égaliseur 23 compense des conditions moyennes de transmission.

Il peut être utilisé seul. Cependant s'il est utilisé seul et qu'une des lignes analogiques est longue, la voix paraît assourdie à la réception. Si au contraire une ligne est très courte, les composantes haute fréquence sont trop présentes. D'autres distorsions du timbre peuvent apparaître systèmes d'émission et de réception ont des réponses fréquentielles trop éloignées des spécifications de l'UIT. C'est pourquoi la pré-égalisation est complétée par un égaliseur adapté, qui adapte la correction de plus précise aux conditions manière réelles de transmission.

L'égaliseur est conçu pour que sa réponse fréquentielle s'adapte automatiquement à la distorsion réelle introduite par la liaison téléphonique en fonction du rapport entre un spectre de référence et le spectre à long terme du signal.

Le principe en est le suivant :

5

10

25

30

Soient s le signal de parole émis par le locuteur, y le signal reçu en bout de chaîne, et h le filtre constitué par le canal analogique complet (émission et réception) et le pré-égaliseur.

D'après la formule des interférences,

$$\gamma_{y}(f) = |H(f)|^{2}. \gamma_{s}(f),$$
 (3)

où γ_s est la densité spectrale de puissance de s, γ_y celle de y et H la réponse fréquentielle de h

Si le canal est supposé invariant dans le temps,

 $E[\gamma_y(f)] = |H(f)|^2. \ E[\gamma_s(f)], \qquad (4), \quad \text{où } E$ 20 désigne la moyenne. Comme $E[\gamma_s(f)]$ n'est pas connu, on l'approche par le spectre moyen de la parole défini par lUIT, que l'on appelle spectre de référence noté $\gamma_{ref}(f)$.

Ainsi on estime la réponse fréquentielle du filtre par :

$$|H(f)|^2 = \frac{E[\gamma_y(f)]}{\gamma_{yy}} \tag{5}$$

La réponse fréquentielle de l'égaliseur adapté a alors pour expression :

$$|EQ(f)| = \sqrt{\frac{\gamma_{ref}(f)}{E[\gamma_{y}(f)]}}$$
 (6)

Comme l'égaliseur est centralisé dans le réseau, γ_y n'est pas connu. On l'exprime en fonction de γ_x densité spectrale de puissance de la sortie x du préégalisateur, dans le cas où il n'y aurait pas d'égalisateur adapté,:

$$\gamma_{y}(f) = \left| L_{RX}(f) \right|^{2} \left| S_{RX}(f) \right|^{2} \cdot \gamma_{x}(f), \tag{7}$$

où L_RX est la réponse fréquentielle de la ligne de réception et S_RX la réponse fréquentielle du système de réception. Comme ces réponses sont inconnues à priori, on les approche par les réponses d'une ligne moyenne et d'un système de réception respectant la spécification de l'UIT, et notées respectivement L_RXo et S_RXo. La réponse fréquentielle de l'égalisateur adaptée recherchée est alors :

$$|EQ(f)| = \frac{1}{|S - RXo(f).L - RXo(f)|} \sqrt{\frac{\gamma_{ref}(f)}{E[\gamma_x(f)]}}, \quad (8)$$

20

25

30

5

On voit dans cette formule que la racine carrée du spectre de référence γ_{ref} est pondéré par le facteur de compensation A(f) du fait de la correction déjà effectuée par le pré-égaliseur. Ce facteur est fonction de la fréquence comme indiqué ci-dessous :

$$A(f) = \frac{1}{|S - RXo(f).L - RXo(f)|}$$

Dans une réalisation préférée, la sortie du préégaliseur 23 est analysée par trames de 32 ms, avec un recouvrement de 50 %.

L'égaliseur 24 adapté est un filtre RIF 251 dont les coefficients sont adaptés à chaque trame d'activité

vocale selon l'équation (8), comme décrit ci-après et représenté sur la figure 7.

Un détecteur de trames d'activité vocale 240 permet de déclencher une chaîne de traitements correspondant aux modules fonctionnels référencés 241 à 251 sur la figure 7 pour obtenir les coefficients du filtre 251.

Une fenêtre d'analyse du signal échantillonné 241 est ouverte. Une transformée de Fourier 242 est appliquée sur les échantillons.

Typiquement, mais non exclusivement, le spectre à long terme de x, $E[\gamma_x]$, est d'abord calculé à partir de l'instant initial d'activité vocale) par moyennage 244 de γ_x sur une fenêtre temporelle croissant de 0 à 4 secondes, puis ajusté récursivement à chaque trame suivante, ce qui se traduit par la formule générique suivante :

$$E[\gamma_{x}(f)]_{n} = \alpha(n)\gamma_{x}(f,n) + (1-\alpha(n))E[\gamma_{x}(f)]_{n-1}$$
(9)

où $E[\gamma_x(f)]_n$ est le spectre à long terme de x à la $n^{\text{ième}}$ trame $\gamma_x(f,n)$ la densité spectrale de puissance de la nième trame, et

$$\alpha(n) = \frac{1}{\min(n,N)}$$

25

5

10

15

où N est le nombre de trames dans 4 secondes.

En pratique γ_x est calculé en prenant le module au carré de la transformée de Fourier rapide 242 sur la figure 7.

30

La réponse fréquentielle de l'égaliseur 24 est donc calculée selon l'équation (8) pour les fréquences comprises entre F_C et F_H , le choix ayant été fait de n'égaliser le signal que sur cette bande.

Les valeurs de |EQ| hors de cette bande de fréquences sont calculées de manière connue par extrapolation linéaire 247 de la valeur en dB de |EQ| (FC-FR)

La réponse impulsionnelle de l'égaliseur est calculée par une transformée de Fourier inverse IFFT 248 de |EQ| suivie d'une symétrisation 249, de manière à obtenir un filtre à phase linéaire.

5

10

15

20

25

30

La réponse fréquentielle de ce filtre, est cependant très irrégulière et, du fait des approximations qui ont entaché son calcul, seule sa forme générale est pertinente.

C'est pourquoi on procède à un étroit fenêtrage symétrique 250 de la réponse impulsionnelle issue des opérations 248 (transformée de Fourier inverse) et 249 (symétrisation). Ce fenêtrage correspond à un lissage de la réponse fréquentielle du filtre.

Le fenêtrage est suivi d'un décalage de manière à obtenir un filtre de la longueur de la fenêtre, sans retard supplémentaire.

On utilise pour cela par exemple une fenêtre triangulaire de longueur 11 (échantillons), dont les coefficients sont représentés sur la figure 6 pour une réponse impulsionnelle initiale sur 256 points. Une fois multipliée par cette fenêtre, la réponse impulsionnelle de l'égaliseur adapté est décalée de 123 points vers la gauche. Ceci permet de ne pas retarder le signal ce qui serait le cas si les zéros devant la fenêtre intervenaient dans le calcul de la sortie du filtre.

On procède alors ensuite à un contrôle automatique de gain. Ce contrôle a typiquement l'un ou l'autre des deux objectifs suivants:

- normaliser le niveau,
- assurer la transparence du dispositif vis-à-vis du niveau global de parole à la réception.

Deux réalisations sont proposées dans la suite correspondant respectivement à ces deux objectifs.

Dans une première réalisation le contrôle de gain est réalisé par l'égaliseur adapté 24. Le choix de γ_{ref} correspond en effet à un niveau nominal souhaitable pour la parole. Ainsi, suivant le niveau choisi pour γ_{ref} , l'égaliseur adapté corrige automatiquement le niveau de parole pour atteindre le niveau souhaité en réception.

Le but de la deuxième réalisation est de corriger le timbre tout en assurant une conservation du niveau global de la parole par rapport à la même liaison sans le dispositif.

Pour cela, le spectre à long terme du signal de réception doit avoir la même énergie avec le dispositif que sans. On applique donc à la sortie de l'égaliseur 24 adapté le gain α défini par la formule théorique :

$$\alpha = \sqrt{\frac{\sum_{k=0}^{255} E[\gamma_{sans} (f)]}{\sum_{k=0}^{255} E[\gamma_{avec} (k)]}}$$
 (10)

25

30

20

5

10

15

avec $\gamma_{avec}(k)$ et $\gamma_{sans}(k)$ densité spectrale de puissance du signal reçu à la fréquence discrète k, respectivement avec et sans l'ensemble pré-égaliseur 23 plus égaliseur adapté 24. Comme le canal est invariant dans le temps, l'égaliseur adapté converge vers une réponse variant peu de sorte qu'à chaque trame :

$$E[\gamma_{avec}(k)] = |EQ_{liss}(k)|^2 |S_RX_0(k)|^2 |L-RX_0(k)|^2 E[\gamma_x(k)]$$
 (11)

où EQ_{liss} est la réponse fréquentielle de l'égaliseur adapté pour la trame courante.

Cependant, $|EQ_{liss}|$ n'est pas connu directement, puisque le lissage de la réponse fréquentielle de l'égaliseur est effectué par un fenêtrage de la réponse impulsionnelle issue de |EQ|. Comme la quantité |EQ| est directement disponible (calculée dans le module d'égalisation adaptée), pour simplifier la réalisation, on approche $|EQ_{liss}|$ par $\lambda |EQ_{liss}|$, λ étant un facteur de correction de la différence d'énergie entre $|EQ_{liss}|$ et |EQ| liée au fenêtrage de la réponse impulsionnelle. Si l'on note W la réponse fréquentielle de la fenêtre,

$$EQ_{liss}(k) = \frac{1}{N} EQ(k) \otimes W(k)$$
 (12)

où \otimes désigne la convolution circulaire et N le nombre de points de la FFT, 256 par exemple. |EQ| étant très irrégulier, si on l'assimile à du bruit,

$$E[|EQ_{liss}(k)|^2] =$$

5

10

15

20

25

$$\frac{1}{N} E[|EQ(k)|^{2}] \sum_{k=0}^{N} W(k)^{2} = E[|EQ(k)|^{2}] \cdot \sum_{n=0}^{N} w(n)^{2} (13)$$

w désignant la fenêtre temporelle. Ainsi,

$$\lambda = \sqrt{\sum_{k=0}^{N-1} w(k)^2} \tag{14}$$

Pour une fenêtre triangulaire de longueur 11, λ =2. D'autre part,

$$E[\gamma_{sans}(k)] = \frac{|S_{RX_{0}}(k)|^{2} |L_{RX_{0}}(k)|^{2}}{|PRE_{EQ}(k)|^{2}} E[\gamma_{x}(k)]$$
(15)

où $\gamma_x(k)$ est la densité spectrale de puissance de la sortie du pré-égaliseur et PRE_EQ(k) la réponse fréquentielle du pré-égaliseur.

Ainsi,

5

10

15

20

25

$$\alpha = \sqrt{\frac{\sum_{k=0}^{255} |S_{RX_{0}}(k)|^{2} |L_{RX_{0}}(k)|^{2}}{|PRE - EQ(k)|^{2}} E[\gamma_{x}(k)]} - (16)$$

$$\sqrt{\sum_{k=0}^{255} \lambda^{2} |EQ(k)|^{2} |S_{RX_{0}}(k)|^{2} |L_{RX_{0}}(k)|^{2} E[\gamma_{x}(k)]}$$

Le gain est donc calculé avec une complexité réduite, puisque $E[\gamma_x(k)]$ et |EQ(k)| sont déjà calculés dans le module d'égalisation adaptée et les autres facteurs sont des constantes.

Si F_c est trop faible, le signal reçu est affecté d'un fort bruit de quantification. En effet, comme l'atténuation du système de réception est d'autant plus forte que la fréquence est faible et que, l'égalisation compensant cette atténuation sur la bande Fc-Fh, est placée avant le système de réception, cette égalisation la sortie du dispositif des induit à anticipée différences de niveau entre les composantes hautes et basses fréquences d'autant plus grandes que Fc est faible. Ainsi, pour certains phonèmes, le niveau du bruit de quantification lors de la conversion en loi A est proche de celui des composantes médium et aiguës. Après l'atténuation des composantes BF par le système 5

10

15

20

25

30

de réception, le bruit perçu à la réception est aussi énergétique que le signal de parole.

Dans une réalisation typique on a choisi :

 F_c = 250Hz, ce qui permet un compromis acceptable entre la restauration des composantes BF et la limitation du bruit de quantification.

Pour tous les locuteurs qui ont été testés, le timbre de la voix à la réception est nettement plus proche de l'original avec la correction par le dispositif présenté que sans dispositif.

Le temps d'adaptation de l'égaliseur est très rapide : une amélioration du timbre est perceptible en moins d'une seconde et une estimation stable de l'égaliseur est obtenue en moins de 4 secondes pour la plupart des locuteurs.

La figure 8 présente pour deux locuteurs la distorsion spectrale entre le signal de parole reçu et le signal original au bout de 4 secondes de parole, dans le cas d'une liaison composée d'un SRI conforme à l'UIT, d'une ligne longue à l'émission et d'une ligne moyenne à la réception.

Une distorsion nulle serait représentée par une courbe plate moyenne (+++) sur toute la bande de fréquences, à -9dB environ si l'on ne modifie pas le niveau de réception par rapport au système sans correction.

La distorsion est représentée dans trois cas : sans correction (tirets), avec pré-égaliseur seul (pointillés fins) et avec le dispositif complet (trait plein). La courbe dans ce dernier cas n'est pas complètement plate sur la bande égalisée (250-3150 Hz), mais les écarts sont inférieurs à 2,5 dB, ce qui est peu perceptible. La correction de niveau est ici

réalisée selon la deuxième méthode (conservation du niveau global).

REVENDICATIONS

1. Procédé dè correction du timbre de la parole transmise sous la forme d'un signal au moyen d'une liaison de transmission d'un réseau téléphonique, la correction ayant lieu après la conversion analogique-numérique du signal dans le réseau, caractérisé en ce qu'il comprend au moins une étape de pré-égalisation du signal numérique par un filtre fixe ayant une réponse fréquentielle dans une bande de fréquences Fc-Fh correspondant à l'inverse d'une déformation spectrale de référence introduite par la liaison téléphonique, avec Fc < 300Hz et Fh ≥ 3150Hz.

5

10

15

20

- 2. Procédé de correction du timbre de la parole selon la revendication 1, caractérisé en ce que la déformation spectrale prise comme référence est caractérisée, pour une liaison du réseau de téléphonie commutée (RTC), à partir de la mise en cascade d'un système de référence intermédiaire (SRI) du type défini par la recommandation UIT-T P.48, et de deux lignes analogiques moyennes (émission et réception).
 - 3. Procédé de correction du timbre de la parole selon la revendication 1, caractérisé en ce que la déformation spectrale prise comme référence est dans le cas où le terminal émetteur est un téléphone mobile et le terminal récepteur est un poste fixe, celle résultant de la mise en cascade d'un filtre respectant le gabarit d'efficacité des mobiles en émission défini par la recommandation P.313 de l'UIT, d'une ligne

10

analogique moyenne et d'un système de récepteur du SRI tel que défini par la recommandation UIT-TP.48.

- 4. Procédé de correction du timbre de la parole selon la revendication 1, caractérisé en ce que la déformation spectrale prise comme référence est, dans le cas où le terminal émetteur est un poste fixe et le terminal récepteur est un téléphone mobile, résultant de la mise en cascade d'un système émetteur du SRI tel que défini par la recommandation UIT-TP.48, analogique ligne moyenne et d'un respectant le gabarit d'efficacité des mobiles réception défini par la recommandation P.313 de l'UIT.
- 5. Procédé de correction du timbre de la parole selon la revendication 1, caractérisé en ce que la déformation spectrale prise comme référence est, dans le cas d'une liaison entre terminaux mobiles, celle résultant de la mise en cascade d'un filtre respectant le gabarit d'efficacité des mobiles en émission défini par la recommandation P.313 de l'UIT et d'un filtre respectant le gabarit d'efficacité des mobiles en réception défini par la recommandation P.313 de l'UIT.
- 6. Procédé de correction du timbre de la parole selon l'une quelconque des revendications précédentes, caractérisé en ce que le filtre utilisé est un filtre de type à réponse impulsionnelle infinie IIR.
- 7. Procédé de correction du timbre de la parole selon l'une quelconque des revendications précédentes, caractérisé en ce que la fréquence basse Fc de la bande de filtrage du filtre est de l'ordre de 250Hz.

8. Procédé de correction du timbre de la parole selon l'une quelconque des revendications précédentes, caractérisé en ce qu'il comporte en outre une étape d'égalisation adaptée dans laquelle on utilise un filtre ayant une réponse fréquentielle adaptée automatiquement à la distorsion réelle introduite par la liaison téléphonique en fonction du rapport entre un spectre de référence et un spectre correspondant au spectre à long terme du signal.

10

15

20

25

- 9. Procédé de correction du timbre de la parole selon la revendication 8, caractérisé en ce que l'étape d'égalisation adaptée comprend :
- la détection d'une activité vocale sur la ligne pour déclencher un enchaînement de traitements pour le calcul des coefficients du filtre numérique en fonction du rapport entre le spectre de référence et le spectre correspondant au spectre à long terme du signal,
- la commande du filtre avec les coefficients obtenus et l'actualisation desdits coefficients,
- le filtrage du signal sortant du pré-égaliseur par ledit filtre.
- 10. Procédé de correction du timbre de la parole selon la revendication 9, caractérisé en ce que l'enchaînement de traitements comprend :
 - le calcul du spectre à long terme du signal dans des fenêtres temporelles successives se recouvrant partiellement,
- le calcul du module de la réponse fréquentielle de l'égaliseur adapté sur la bande Fc-Fh en effectuant le rapport de la racine carrée du spectre à long terme obtenu dans une fenêtre temporelle à la racine carrée du spectre de référence, la racine carrée du spectre de

référence étant compensé à chaque fréquence par un facteur prédéterminé A(f).

11. Procédé de correction du timbre de la parole selon la revendication 9 ou 8, caractérisé en ce que l'enchaînement de traitements comprend une extrapolation du module de la réponse fréquentielle de l'égaliseur adapté, pour les fréquences en dehors de la bande Fc-Fh, typiquement pour une bande 0-4000Hz.

10

15

- 12. Procédé de correction du timbre de la parole selon l'une quelconque des revendications 9 à 11, caractérisé en ce que l'enchaînement de traitements comprend le calcul de la réponse impulsionnelle du filtre numérique à partir du module de la réponse fréquentielle de l'égaliseur adapté extrapolé pour les fréquences en dehors de la bande Fc-Fh.
- 13. Procédé de correction du timbre de la parole selon l'une quelconque des revendications 8 à 12, caractérisé en ce que le calcul du spectre du signal à long terme comprend une opération de transformée de Fourier rapide.
- 14. Procédé de correction du timbre de la parole 25 selon l'une quelconque des revendications 8 à caractérisé que le calcul de la en ce. impulsionnelle du filtre consiste à calculer coefficients du filtre en opérant une transformée de réponse fréquentielle inverse sur la Fourier 30 symétrisation puis une l'égaliseur adapté, une opération, assurant le lissage de la réponse fréquentielle.

15. Procédé de correction du timbre de la parole selon la revendication 14, caractérisé en ce que l'opération de lissage est effectuée par l'application d'une fenêtre temporelle sur la réponse impulsionnelle.

5

10

15

25

- 16. Procédé de correction du timbre de la parole selon l'une quelconque des revendications 8 à 15, caractérisé en ce que le calcul du spectre à long terme du signal dans des fenêtres temporelles successives se recouvrant partiellement comprend :
 - l'échantillonnage du signal dans une fenêtre temporelle,
 - l'opération de transformée de Fourier rapide (FFT) du signal échantillonné,
 - le calcul de la densité spectrale de puissance,
 - le calcul de la moyenne de la densité spectrale de puissance, sur une durée prédéterminée.
- 17. Procédé de correction du timbre de la parole 20 selon la revendication 16, caractérisé en ce que le calcul de la moyenne de la densité spectrale de puissance consiste:
 - pour les N premières fenêtres temporelles à partir de la détection de présence de parole dans le signal, calculer la arithmétique des densités spectrales puissance de toutes les fenêtres temporelles écoulées depuis ladite détection, N étant un nombre de fenêtre temporelles prédéterminé, typiquement mais non exclusivement le nombre de fenêtres temporelles dans 4 secondes de parole ;
 - pour les fenêtres temporelles suivantes, à ajuster la moyenne de la densité spectrale de puissance calculée à la fenêtre temporelle

précédente par un lissage récursif du premier ordre tenant compte de la densité spectrale de puissance de la fenêtre temporelle courante.

- 18. Procédé de correction du timbre de la parole selon l'une quelconque des revendications 8 à 17, caractérisé en ce qu'il comprend une étape de contrôle automatique du gain.
- 10 19. Procédé de correction du timbre de la parole selon la revendication 18, caractérisé en ce que le contrôle automatique du gain est réalisé l'enchaînement des traitements de l'étape d'égalisation adaptée en choisissant une densité spectrale 15 référence Yref correspondant au niveau souhaité en réception.

20

- 20. Procédé de correction du timbre de la parole selon la revendication 18, caractérisé en ce que le contrôle automatique du qain est réalisé par amplification du signal obtenu après égalisation adaptée avec un gain α fonction du rapport entre la densité spectrale de puissance du signal de réception lorsque l'on a effectué une pré-égalisation et une égalisation adaptée du signal et, la densité spectrale de puissance du signal en l'absence de pré-égalisation et d'égalisation du signal.
- 21. Filtre numérique fixe destiné à la correction 30 du timbre de la parole dans un réseau de transmission téléphonique, caractérisé en ce qu'il a une réponse fréquentielle dans une bande de fréquences Fc-Fh correspondant à l'inverse d'une déformation spectrale

de référence introduite par la liaison téléphonique, avec Fc < 300 Hz et Fh $\geq 3150 \text{Hz}$.

- 22. Filtre numérique adapté destiné à la correction du timbre de la parole dans un réseau de transmission téléphonique, caractérisé en ce qu'il comprend des moyens de traitement du signal de parole ayant une réponse fréquentielle adaptée automatiquement à la distorsion réelle introduite par la liaison téléphonique en fonction du rapport entre un spectre de référence et un spectre correspondant au spectre à long terme du signal.
- 23. Dispositif de correction du timbre de la parole dans un réseau de transmission téléphonique, caractérisé en ce qu'il comprend un filtre fixe selon la revendication 20 suivi d'un filtre adapté selon la revendication 21 et des moyens de contrôle automatique du gain.

20

5

10

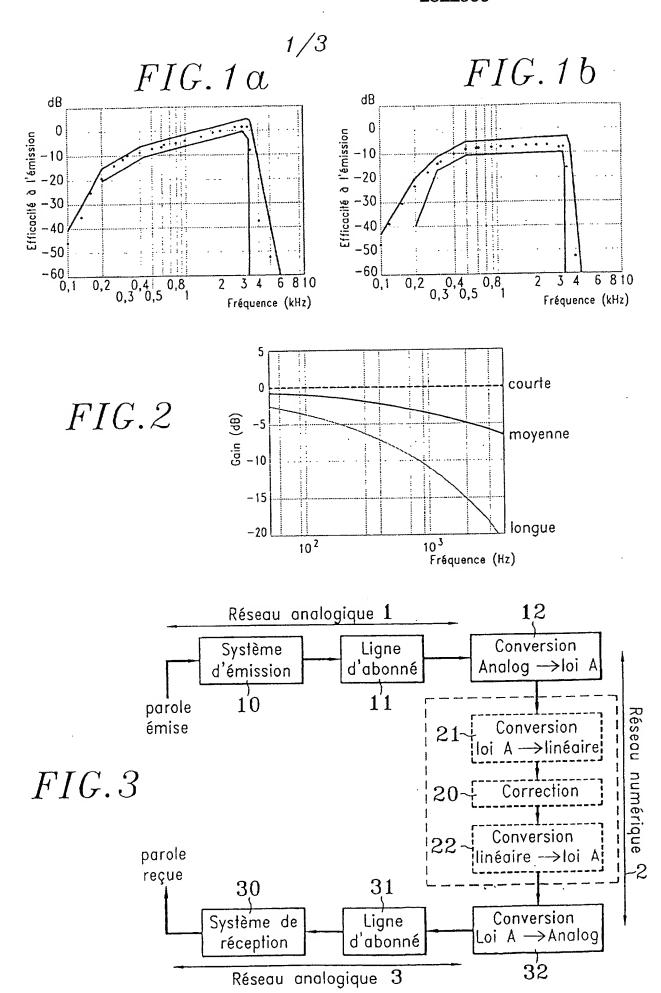
24. Dispositif de correction du timbre de la parole dans un réseau de transmission selon la revendication 23, caractérisé en ce que les moyens de contrôle automatique du gain sont réalisés par le filtre adapté.

25

30

25. Dispositif de correction du timbre de la parole dans un réseau de transmission selon la revendication 23, caractérisé en ce que les moyens de contrôle automatique du gain sont réalisés un amplificateur du signal de gain α fonction du rapport entre la densité spectrale de puissance du signal de réception lorsque l'on a effectué une pré-égalisation et une égalisation adaptée du signal et, la densité spectrale de puissance

du signal en l'absence de pré-égalisation et d'égalisation adaptée du signal.



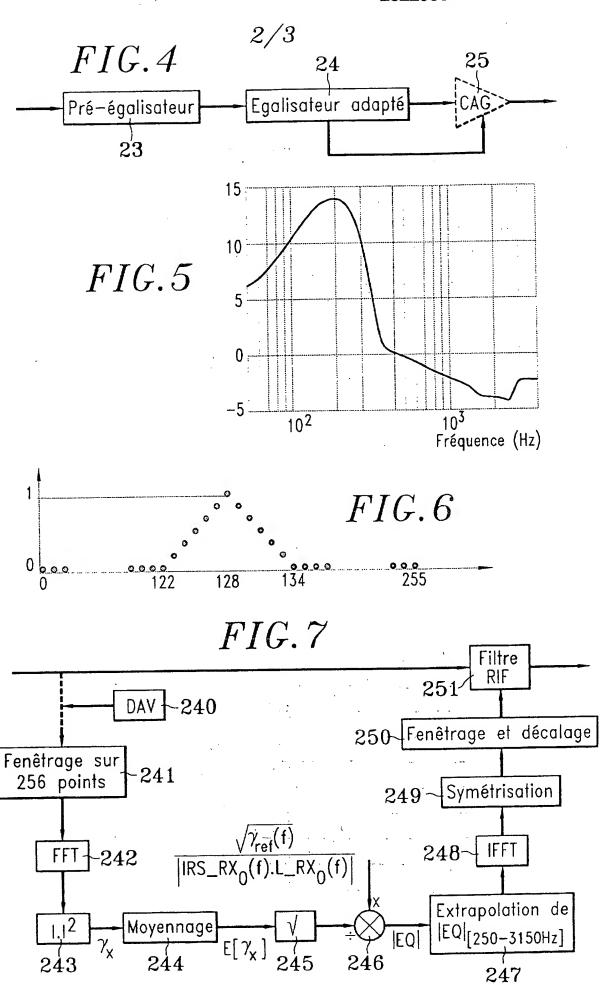
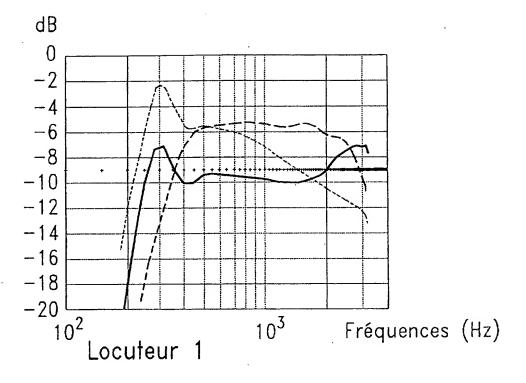
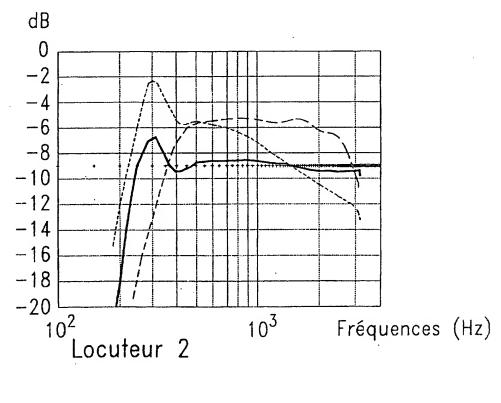


FIG.8







RAPPORT DE RECHERCHE **PRÉLIMINAIRE**

N° d'enregistrement national

établi sur la base des dernières revendications déposées avant le commencement de la recherche FA 600963 FR 0104194

DOCU	MENTS CONSIDÉRÉS COMME PERTINENTS	Revendication(s) concernée(s)	Classement attribué à l'invention par l'INPI
atégorie	Citation du document avec indication, en cas de besoin, des parties pertinentes		. i
(US 5 896 449 A (OSHIDARI ET AL) 20 avril 1999 (1999-04-20)	1,2, 7-11,13, 16,18-25	G10L21/02 H04M3/40 H04M9/00
ľ	<pre>* abrégé; figures 1,2,6 * * colonne 6, ligne 36-56 *</pre>	3-5	H04M1/253 H04L27/01 H03G3/20
t .	DE 198 52 091 C (DEUTSCHE TELEKOM MOBIL) 25 mai 2000 (2000-05-25) * abrégé *	3-5	H03G5/22
Ą	US 6 157 909 A (MAUUARY ET AL) 5 décembre 2000 (2000-12-05) * abrégé; figures 2A-B *	1-25	
D,A	US 5 915 235 A (DEJACO ET AL) 22 juin 1999 (1999-06-22) * abrégé; figures 1,4,5 *	1-25	
		·	
			DOMAINES TECHNIQUES RECHERCHÉS (Int.CL.7)
			G10L H03G
	-		
		esse come a	y
	Date d'achèvement de la recherche		Examinateur
	31 octobre 200	ı Qu	élavoine, R

1

EPO FORM 1

A: arrière-plan technologique
O: divulgation non-écrite
P: document intercalaire

D : cité dans la demande L : cité pour d'autres raisons

&: membre de la même famille, document correspondant

ANNEXE AU RAPPORT DE RECHERCHE PRÉLIMINAIRE RELATIF A LA DEMANDE DE BREVET FRANÇAIS NO. FR 0104194 FA 600963

La présente annexe indique les membres de la famille de brevets relatifs aux documents brevets cités dans le rapport de recherche préliminaire visé ci-dessus. Les dits membres sont contenus au fichier informatique de l'Office européen des brevets à la date d31-10-2001 Les renseignements fournis sont donnés à titre indicatif et n'engagent pas la responsabilité de l'Office européen des brevets, ni de l'Administration française

Document brevet cité au rapport de recherche		Date de publication		Membre(s) de la famille de brevet(s)		Date de publication	
US 589	96449	Α	20-04-1999	US	5592545	A	07-01-1997
00 00.		••		US	5471527		28-11-1995
				ΑÜ	708752		12-08-1999
				AU	1293697		28-07-1997
				BR	9612389		28-12-1999
				CA	2241903		10-07-1997
				CN	1212100		24-03-1999
				EP	0870391		14-10-1998
				JР	•	T	07-03-2000
				PL	329229	A1	15-03-1999
				WO	9724858		10-07-1997
				AU	690603	B2	30-04-1998
	•			AU	1101895	Α	19-06-1995
				AU	6482298	Α	02-07-1998
				BR	9408202	Α	26-08-1997
				CA	2177266	A1	08-06-1995
				CN	1149373	Α	07-05-1997
				EP	0732024	A1	18-09-1996
			•	FI	962305		31-05-1996
				JP	3032811	B2	17-04-2000
				JP	9506220	T	17-06-1997
				KR	268319	B1	16-10-2000
				PL	314880		30-09-1996
				RU	2142675		10-12-1999
				WO	9515643		08-06-1995
				ZA	9409489	Α	14-08-1995
DE 19	852091	С	25-05-2000	DE	19852091		25-05-2000
				AU	1963200		05-06-2000
				WO	0030277		25-05-2000
				EP	1145575	A2	17-10-2001
US 61	57909	A	05-12-2000	FR	2766604	A1	29-01-1999
				AU	8813398		16-02-1999
				BR	9806232	Α	21-03-2000
				CN	1234935		10-11-1999
				EP	0932964		04-08-1999
				WO	9905831		04-02-1999
				JP	2001501327	T 	30-01-2001
US 59	1 52 35	A	22-06-1999	AUCI	JN		

THIS PAGE BLANK (USPTO)